

**USDA Service Center Initiative
Geospatial Data Acquisition, Integration and Delivery
Business Re-engineering Project**

Data Themes - Outline - GNIS Cities

I. Acquisition

A. Data Source

1. Producer Information

a. Name

The U.S. Populated Places File is a thematic subset of the Geographic Names Information System (GNIS), produced by the United States Geological Survey (USGS).

b. Location of Headquarters

US Geological Survey
523 National Center
Reston, VA 20192 USA

c. Internet Address

www.usgs.gov

2. Publisher Information

a. Name

The U.S. Populated Places File is published by the USGS. It is available through the National Geospatial Data Clearinghouse.

b. Location of Headquarters

Branch of Geographic Names
US Geological Survey
523 National Center
Reston, VA 20192 USA

c. Internet Address

The internet address of the USGS is www.usgs.gov
The internet address of the clearinghouse is www.nsd.usgs.gov

3. Acquisition Information

a. Delivery Media

The U.S. Populated Places file is available via anonymous file transfer protocol (ftp). It is also available on CD-ROM, 9-track magnetic tape or 8-mm cartridge in ASCII code, and computer generated listings are available on bound paper.

b. Download URL

<http://mapping.usgs.gov/www/gnis/gnisftp.html>

c. Projected Data Availability Schedule

The U.S. Populated Places file is available now. The file lists information about all cities and towns throughout the United States that are described in the Geographic Names Information System (GNIS) database. The GNIS database is being compiled in two phases. The first phase is complete for all States and Territories, and captured feature names from USGS large-scale topographic maps, U.S. Forest Service maps, National Ocean Service charts, Federal Aviation Administration files, Federal Communications Commission files, and files of the Army Corps of Engineers.

The second phase of compilation is complete or in progress for about 65 percent of the United States, and captures names from State, locally, and privately published current and historical maps, charts, and texts. The status of States and territories for phase II is available at <http://mapping.usgs.gov/pub/gnis/00README.html>. The status at the time of writing is listed below:

Complete	In progress	Not Started
Alabama	California	Alaska
Arizona	Illinois	American Samoa
Arkansas	Iowa	Connecticut
Colorado	Massachusetts	Guam
Delaware	Minnesota	Hawaii
District of Columbia	North Carolina	Kentucky
Florida	Oklahoma	Maine
Georgia	Wisconsin	Michigan
Idaho		Micronesia
Indiana		New Hampshire
Kansas		Puerto Rico
Louisiana		Rhode Island
Maryland		Texas
Mississippi		Vermont
Missouri		Virgin Islands
Montana		
Nebraska		
Nevada		
New Jersey		
New Mexico		
North Dakota		
Ohio		
Oregon		
Pennsylvania		
South Carolina		
South Dakota		
Tennessee		
Utah		
Virginia		
Washington		
West Virginia		
Wyoming		

B. Standards Information

1. Geospatial Data Standard

a. Standard Name and Steward Information

None. Data is tabular. Coordinates are listed as degrees, minutes, and seconds of longitude and latitude.

b. Standard Version

None. Data is tabular.

c. Standard URL

None. Data is tabular.

2. Metadata Standard

a. Standard Name and Steward Information

Content Standards for Digital Geospatial Metadata version 19940608

The metadata contact is:

US Geological Survey
508 National Center
Reston, VA 22092 USA

b. Description of Metadata Captured

The metadata is available online at <http://nsdi.usgs.gov/wais/maps/gnis.html>

The sections of metadata include:

Identification Information
Data Quality Information
Spatial Data Organization Information
Spatial Reference Information
Entity and Attribute Information
Distribution Information
Metadata Reference Information

c. Metadata Accuracy and Completeness Assessment

The metadata is complete. Additional information is available in the

U.S. Department of the Interior, U.S. Geological Survey, 1987 Geographic Names Information System--Data Users Guide 6; Reston, Virginia Softcopy in hypertext format is available at:

http://mapping.usgs.gov/www/ti/GNIS/gnis_users_guide_toc.html

C. Acquired Data Structure

1. Geospatial Data Format

a. Format (raster, vector, etc.)

The U.S. Populated Place data is acquired as an ASCII, tabular file. The file has been compressed with the GNU gzip utility. If you do not have access to gzip, the FTP server will uncompress the file as you retrieve it.

b. Format Name

ASCII

c. Data Extent

The United States and its territories. Covers -180, -12 degrees SW to 172, 72 degrees NE.

d. Horizontal and Vertical Resolution

Horizontal resolution is 1 decimal second.

Vertical resolution is 1 foot.

e. Absolute Horizontal and Vertical Accuracy

Accuracy of these digital data is based upon the use of source graphics that are compiled to meet National Map Accuracy Standards. Comparison to the graphic source is used as control to assess digital positional accuracy.

f. Nominal Scale

1:24,000

g. Horizontal and Vertical Datum

The horizontal datum is the North American Datum (NAD) 83. The vertical datum is mean sea level.

h. Projection

None

i. Coordinate Units

Degrees, minutes, and decimal seconds

j. Average Data Set Size

The U.S. Populated Places file is 3.9 Mb in size when acquired in gzip format. It covers the entire U.S. and its territories. When uncompressed, the ASCII file is 26.6 Mb.

k. Symbology

None

2. Attribute Data Format

a. Format Name

Fixed format ASCII

b. Database Size

The U.S. Populated Places database is 26.2 Mb uncompressed, 3.9 Mb in gzip format.

3. Data Model

a. Geospatial Data Structure

None

b. Attribute Data Structure

See below.

c. Database Table Definition

Populated Places

Columns	Contents
1 - 49	Feature Name
50 - 54	Feature Type
55 - 58	State
59 - 78	County
79 - 82	State FIPS Code
83 - 88	County FIPS Code
89 - 107	Geographic Coordinates
108-117	Federal Status of Feature Name
118-125	Feature Elevation
126-148	Topographic Map
149-157	Population

d. Data Relationship Definition

The U.S. Populated places file is a list of all of the cities and towns from the GNIS database.

e. Data Dictionary

Feature Name

The Federally recognized name of the physical or cultural feature.

Feature Type

ppl - (populated place) place or area with clustered or scattered buildings and a permanent human population (city, settlement, town, village).

County

County name

State/County FIPS Code

Combination of the 2 character state FIPS code and 3 character county FIPS code. Includes leading zeros. For example, 08069 is Larimer County, Colorado.

Geographic Coordinates

Coordinates locate the approximate center of an areal feature or the mouth of a linear feature. Coordinate units are degrees, minutes, and seconds, ranging from 180W to 180E longitude, and 90S to 90N latitude.

Federal Status

Federal status of the feature name. A variable -length alphabetic entry indicates the year of any decision by the U.S. Board on Geographic Names regarding the

feature name or its application. This means the name or application of the name of the feature has been the subject of controversy and has undergone special research and consideration by the Board. The format of the entry is BGN YEAR, where YEAR is the appropriate year of the BGN decision. If the name has been made official by an act of Congress, the format will be US YEAR.

Feature Elevation

Elevation of the feature in feet above mean sea level, listed when available.

Topographic Map

The name of the USGS 1:24,000 topographic map on which the feature is located is listed.

Population

Estimated 1994 population of incorporated cities and towns

D. Policies

1. Restrictions

a. Use Constraints

None. Some of the data provided were derived from sources outside of the U.S. Geological Survey. All data and information contained in any of the data files are provided without guarantee as to their completeness or correctness. Any conclusions drawn from these data and information are the sole responsibility of the user.

b. Access Constraints

None.

c. Certification Issues

None.

2. Maintenance

a. Temporal Information

Database is re-published monthly with additional data.

b. Average Update Cycle

Unknown. Documented as “as needed” in the metadata.

E. Acquisition Cost

1. Cooperative Agreement

a. Description of Agreement

None

b. Status of Agreement

N/A

2. Cost to Acquire Data

None, assuming it is acquired via ftp.

II. Integration

A. *Value Added Process*

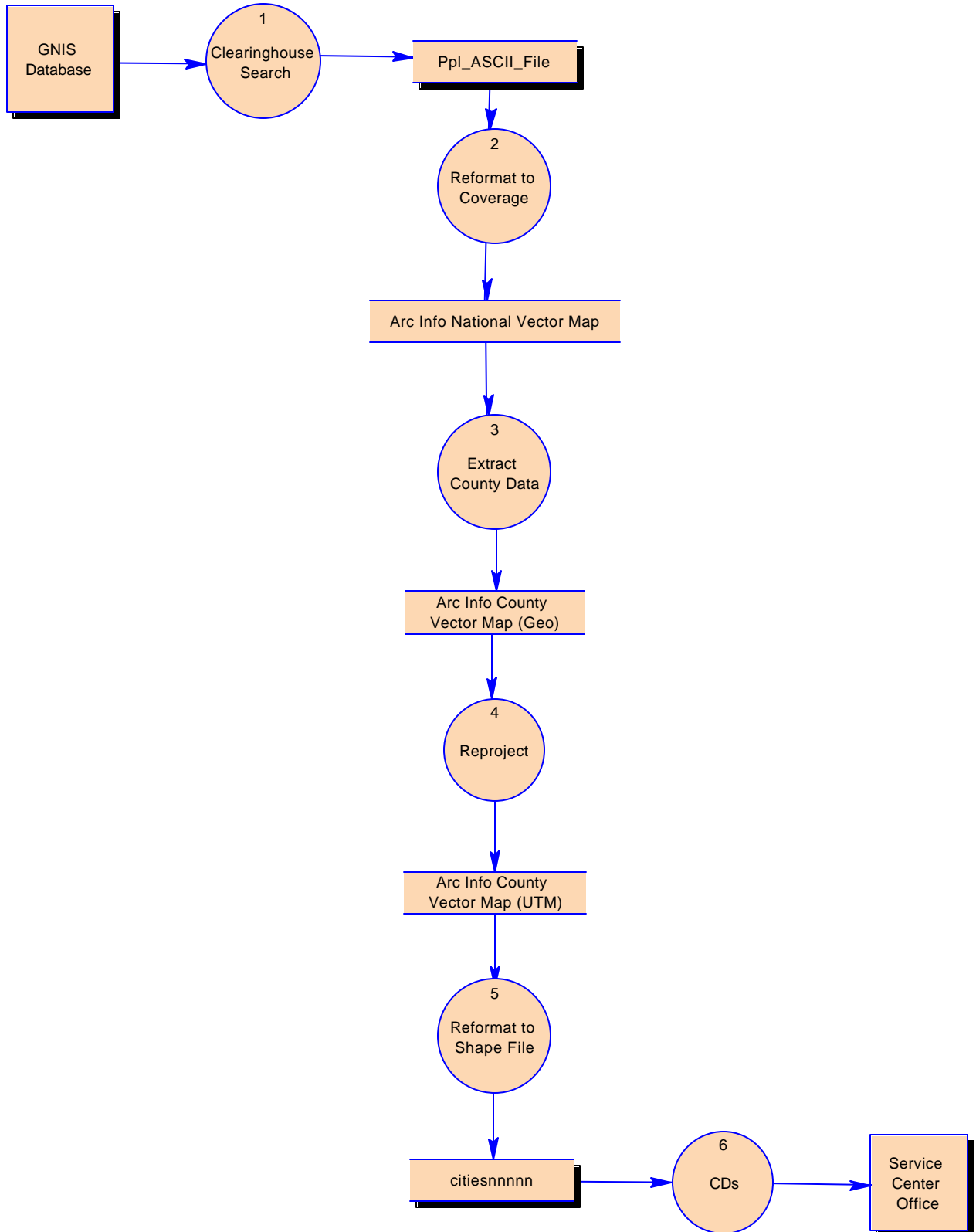
1. Benefit to the Service Center

The U.S. Populated Places ASCII, tabular data is converted to a vector map. Without this, the cities data cannot be displayed or analyzed with other geospatial data. If a Service Center wanted to use the populated place data, it would have to perform the conversion. In addition, the cities data is tiled by county and reprojected to the appropriate UTM zone. Without these processes, the service center would have to remove unwanted data and reproject the data from geographic if they wanted to view or analyze the data with any other UTM spatial data.

2. Process Model

a. Flow Diagram

Process Model		
Project : Data AID		
Model : City Process Model		
Author : Jean Maloney	Version 1.0	12/21/98



b. Process Description

- The U.S. Populated Places ASCII file is downloaded from the USGS ftp site
- The U.S. Populated Places ASCII file is converted to an Arc/Info vector map
- The Arc/Info vector map is converted to a shape file.
- The populated places points which are located within the specified county are selected and saved as a separate shape file
- The county shape file is reprojected from geographic (lat/long) to the correct UTM zone for that area.

3. Technical Issues

a. Tiling

Original data is in a single tile for the entire nation and its territories. This is changed to a county tiling scheme. The only time this may be a problem would be when the data in multiple counties is projected into different UTM zones and the entire dataset must be analyzed.

b. Compression

None.

c. Scale

The populated place locations are taken from 1:24,000 scale maps. Therefore, the data should not be used at a scale larger than this.

d. Tonal Matching

Not applicable.

e. Edge-matching

Not applicable. Information is point data. When a point is located exactly on a county boundary, it will be located at the identical coordinates on the adjacent county. An exception might arise when adjacent counties are located in different UTM zones.

4. Quality Control

a. Procedures

A random sample of 10% of the entries in the system were visually verified against the compilation source data (large-scale USGS topographic maps) to ensure an accuracy rate of at least 95. Locative references (geographic coordinates, topographic map, and county) are cross-checked for logical consistency.

b. Acceptance Criteria

See above.

5. Data Steward

a. Name and Organization

Currently, the data steward for the integrated cities data is:

National Cartography and Geospatial Center
Natural Resources Conservation Service
US Department of Agriculture
501 Felix Street, Building 23
P. O. Box 6567
Fort Worth, Texas 76115-0567 USA

If the integration procedure can be automated, the steward would optimally remain:

Roger L. Payne
US Geological Survey
523 National Center
Reston, VA 20192 USA

b. Responsibilities

The USGS National Mapping Program maintains the system. Each regional mapping center of the USGS, as well as the U.S. Forest Service and National Ocean Service, compiles and electronically transmits names data to the GNIS staff for review and entry into the system. The U.S. Board on Geographic Names transmits information directly to the system concerning new names and the resolution of geographic names that are in conflict on Federal sources. A series of checks and balances ensures integrity and security so that all users can retrieve and use data with confidence.

If the responsibility of the cities data set remains that of USDA, we will be responsible for updating the USDA version of the data on a periodic basis. This includes re-integrating the data.

B. Integrated Data Structure

1. Geospatial Data Format

a. Format (raster, vector, etc.)

Vector

b. Format Name

ESRI shape file and Arc/Info coverage

c. Data Extent

Individual county

d. Horizontal and Vertical Resolution

Same as source data.

e. Absolute Horizontal and Vertical Accuracy

Same as source data.

f. Nominal Scale

Same as source data.

g. Horizontal and Vertical Datum

The horizontal datum is the North American Datum (NAD) 83. The vertical datum is mean sea level.

h. Projection

Universal Transverse Mercator (UTM), North American Datum (NAD) 83

i. Coordinate Units

Meters

j. Symbology

Hollow circle with a black outline. The diameter is dependent on the population field of the city attributes.

2. Attribute Data Format

a. Format Name

Dbase V, as part of an ESRI Shape file

b. Database Size

The data per county ranges from approximately 5 to 1000 records

3. Data Model

a. Geospatial Data Structure

Shape Files	
map shp	shp file
map dbf	dbf file
map shx	shx file
map sbn	sbn file
map sbx	sbx file

b. Attribute Data Structure

See below.

c. Database Table Definition

Cities_pnt	
<u>Cities_id</u>	I
Shape	N11,5
Area	N11,5
Perimeter	N11,5
Cities_	I
Name	A100
Type	A9
State	A2
County	A35
Latitude	N11,5
Longitude	A8
Status	A10
Elevation	A6
Topo_map	A54
Population	I
Symbol	Symbol
Dx	A3
Mx	A2
Sx	A2
Dy	A3
Mv	A2

d. Data Relationship Definition

The cities data is a subset of the GNIS data theme. It is all of the populated places.

e. Data Dictionary

Cities_id

A unique identifier for each feature. A remnant of Arc/Info data processing. Not required.

Shape

A mandatory field for all shape files. Will appear as the string point for each point feature.

Area

The area of polygon features. A remnant of Arc/Info processing. Not required. Always the value 0.0 because the data are point features.

Perimeter

The perimeter of polygon features or length of linear features. A remnant of Arc/Info processing. Not required. Always the value 0.0 because the data are point features.

Cities_

A unique identifier for each feature. A remnant of Arc/Info data processing. Not required.

Name

The Federally recognized name of the physical or cultural feature.

Type

One of the types of features listed above under Source Data.

County

County name

Fips

Combination of the 2 character state FIPS code and 3 character county FIPS code. Includes leading zeros. For example, 08069 is Larimer County, Colorado.

Latitude

Coordinates locate the approximate center of populated place. Coordinate units are degrees, minutes, and seconds, ranging from 90S to 90N latitude.

Longitude

Coordinates locate the approximate center of populated place. Coordinate units are degrees, minutes, and seconds, ranging from 180W to 180E longitude.

Status

Federal status of the geographic name.

Elevation

Elevation of the feature in feet above mean sea level, listed when available.

Topo_map

The name of the USGS 1:24,000 topographic map on which the feature is located is listed.

Population

Human population of the populated place.

Symbol

Symbol to use.

Dx

The degrees of the feature's longitude.

Mx

The minutes of the feature's longitude.

Sx

Dy	The seconds of the feature's longitude
My	The degrees of the feature's latitude.
Sy	The minutes of the feature's latitude.
Sy	The seconds of the feature's latitude.

C. Resource Requirements

1. Hardware and Software

To acquire the populated place data, it requires a UNIX or NT machine with approximately 50 Mb of disk.

2. Staffing

It requires one staff member for approximately ½ hour to acquire the entire populated place database from the USGS ftp site.

D. Integration Cost

1. Hardware and Software

In order to reformat, reproject, and tile the data, the USDA requires:
Arc/Info on UNIX or NT platform
ArcView on NT platform
1 GB disk

2. Staffing

The procedure is currently not 100% automated. If it was automated, this procedure would require personnel only to check the results of the procedure. A rough estimate for fully automating the procedure is one programming staff member for 5 days. The automated procedure would require approximately one day to run.

To generate maps for the US and territories with no automation, it would require approximately 2 staff members, familiar with ESRI software, for 5 days.

III. Delivery

A. Specifications

1. Directory Structure

- a. Folder Theme Data is Stored In

\Government Units (V 6.0)

2. File Naming Convention

- a. List of Theme Files and The File Naming Convention (V6.0, non 8.3 compliant)

\citiesnnnnn.dbf
\citiesnnnnn.shx
\citiesnnnnn.shp

where nnnnn = <StFIPS><CtyFIPS>, e.g. cities18143.shp from State FIPS 18, County FIPS 143.

B. User Information

1. Accuracy Assessment

a. Alignment with Other Theme Geospatial Data

Will be 100% aligned with the GNIS data theme. The cities will be aligned well with other themes captured at the source scale for the populated places. It should be noted by the data user that the largest scale maps available were used during Phase I compilation and the majority of the names were compiled from the 1:24,000-scale, 7.5-minute USGS topographic maps. For areas where no published or advanced 1:24,000-scale maps existed, the 1:62,500-scale maps were used; where there was no coverage by either series of maps, the 1:250,000-scale maps were used. Therefore, the positional accuracy of the cities data will vary, depending on the maps available at the location of each city.

b. Content

Because the U.S. Board on Geographic Names determines the city names, they are the most standard names available in the U.S. The feature elevation and populations are only given when known.

2. Appropriate Uses of the Geospatial Data

a. Display Scale

The original data source scale or smaller, usually 1:24,000.

b. Plot Scale

The original data source scale or smaller, usually 1:24,000.

c. Area Calculations

Not applicable. Cities data is point data.

d. Decision Making

The city locations are the approximate centroids of area features. The coordinates should not be used only for approximate calculations.

C. Maintenance and Updating

1. Recommendations and Guidelines

a. Frequency of Updates

In order of preference:

- Extract the data from the USGS clearinghouse node at the time of request for the data. Perform the data integration in an automated fashion. Therefore, no updates are required because USDA would not be the data steward.
- To coincide with USGS updates, if notification is a possible from USGS
- At a regular interval of 3 months, 6 months or 12 months, depending on budget

b. Location for the Theme Data to be Maintained

In order of preference:

- At the USGS, with USGS as the data steward
- At the USDA data warehouse, potentially in Fort Worth

c. Maintenance and Updating Procedures Overview

Follow the integration procedure listed above for each update if it is not done at the time of request of the data.